Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 79 (2016) 1029 – 1035

Procedia Computer Science

7th International Conference on Communication, Computing and Virtualization 2016

Client side Privacy Protection Using Personalized Web Search

Mrs. Sharvari V. Malthankar , Prof. Shilpa Kolte

PG Student,Saraswati College of Engg.Kharghar,NaviMumbai,India.
Assistant Professor, Saraswati College of Engg.Kharghar,NaviMumbai,India.

Abstract

We are providing a Client-side privacy protection for personalized web search.. Any PWS captures user profiles in a hierarchical taxonomy. The system is performing online generalization on user profiles to protect the personal privacy without compromising the search quality and attempt to improve the search quality with the personalization utility of the user profile. On other side they need to hide the privacy contents existing in the user profile to place the privacy risk under control. User privacy can be provided in form of protection like without compromising the personalized search quality. In general we are working for a trade off between the search quality and the level of privacy protection achieved from generalization.

Peer-review under responsibility of the Organizing Committee of ICCCV 2016

Keywords: UPS,Privacy Protection, Greedy DP,Greedy IL;

1. Introduction

The web search engine has long become the most important portal for ordinary people looking for useful information on the web. users may experience failure when search engines return irrelevant results that do not meet their real and expected intentions. Such irrelevant think is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of the texts. Personalized web search (PWS) is one general search techniques aiming to providing better search results, which are tailored to individual user needs. At the expense, user information has to be collected and analyzed to figure out the user intention behind the issued query.

PWS can generally into two types

- Click-log-based methods and
- Profile-based methods

Peer-review under responsibility of the Organizing Committee of ICCCV 2016
doi:10.1016/j.procs.2016.03.130

In Click-log-based methods we found as

- They simply impose bias to clicked pages in the user's query history.
- It can only work on repeated queries from the same user, which is a strong limitation confining its applicability.

In Profile-based methods we found as

- Profile-based methods can be potentially effective forAlmost all sorts of queries, but are reported to be Unstable under some circumstances.
- Improve the search experience with complicated user-interest models generated from user profiling techniques.
- PWS has demonstrated more effective in improving the quality of web search recently, with increasing usage of personal and behavior information to profile its users, which is usually gathered implicitly from query history, browsing history, click-through data bookmarks, user documents and so forth.

2 LITERATURE SURVEY

Profile-Based Personalization

Many profile representations are available in the literature to facilitate different personalization strategies. Earlier techniques utilize term lists/vectors or bag of words to represent their profile. However, most recent works build profiles in hierarchical structures due to their stronger descriptive ability, better scalability, and higher access efficiency. The majority of the hierarchical representations is constructed with existing weighted topic hierarchy/graph, such as ODP, Wikipedia and so on. Another work builds the hierarchical profile automatically via term-frequency analysis on the user data.

Privacy Protection, in PWS System

- Generally there are two classes of privacy protection problems for PWS.
- One class includes those treat privacy as the identification of an individual.
- The author includes those consider the sensitivity of the data, particularly the user profiles, exposed to the PWS server.
- The useless user profile (UUP) protocol is proposed to shuffle queries among a group of users who issue them. As a result any entity cannot profile a certain individual. These works assume the existence of a trustworthy third-party anonymizer, which is not
  Readily available over the Internet at large.

- [1] And [2] We propose a PWS called UPS that can adapt generalize profiles by queries while respecting user-specified privacy requirements. Our run time generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile.

- Propose a method that, given a query submitted to a search engine, suggests a list of related queries The related queries are based in previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process The method proposed is based on a query clustering process in which groups of semantically similar queries are identified The clustering process uses the content of historical preferences of users registered in the query log of the search engine The method not only discovers the related queries, but also ranks them according to a relevance criterion Finally, we show with experiments over the query log of a search engine the effectiveness of the method [3].

- Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various

search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier to the wide proliferation of PWS [4].

- We study a large-scale evaluation framework for personalized search based on query logs and then evaluate five personalized search algorithms (including two clicks-based ones and three topical-interest-based ones) using 12-day query logs of Windows Live Search. By analyzing the results, we reveal that personalized Web search does not work equally well under various situations [5].

- Long-term search history contains rich information about a user's search preferences, which can be used as search context to improve retrieval performance. The user profiles for particular users are stored on the clients, thus preserving the privacy of the users. The design adopts the server-client model in which user queries are forwarded to a server for processing the training and re-ranking quickly [6].

- The proposed introduces vector quantization approach piecewise on the datasets which segment, each row of datasets and quantization approach is performed on each segment, using the proposed approach which later are again united to transformed data set [7].

- We study private safety in pws applications that representation user desire as hierarchical user profiles. We are providing a private requirement using a pws framework ups. Two predictive metrics utility of personalization and the privacy risk are used for build – up of the profile. In the generalization process we use greedy DP and the greedy IL algorithm. The innovative outcome tells that greedy IL obviously outperforms greedy DP in terms of efficiency [8].

- We propose a method that, given a query submitted to a search engine, suggests a list of related queries. The related queries are based on previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process. The method proposed is based on a query clustering process in a group of semantically similar queries are identified. [9].
- We proposed the reliability of implicit feedback generated from click through data in WWW search. Analyzing the users' decision process using eye tracking and comparing implicit feedback against manual relevance judgments, we conclude that clicks are informative but biased. We show that relative preferences derived from clicks are reasonably accurate on average [10].

- We propose a novel context-aware query suggestion approach.. In which steps for in an offline model learning step, to address data sparseness, queries are summarized into concepts by clustering a click-through bipartite. Then, from session data a concept sequence suffix tree is constructed as the query suggestion model [11].

3. PROBLEM DEFINITION

To protect user privacy in profile-based PWS, we have to consider two contradicting effects during the search process. On the one hand, they attempt to improve the search quality with the personalization utility of the user profile. They need to hide the privacy contents in existing user profile to place the privacy risk under control. Significant gains can be obtained by personalization at the expense of only a small and less-sensitive portion of the user profile, namely a generalized profile. Thus, user privacy can be protected without compromising the personalized search quality

The existing profile-based Personalized Web Search does not support runtime profiling. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly. Such "one profile is fits all" strategy certainly has drawbacks given the variety of queries. One evidence reported in is that profile-based personalization may not even help to improve the search quality for some ad hoc queries, though exposing user profile to a server has put the user's privacy at risk.

The existing methods do not take into account of the customization of privacy requirements. This makes some user privacy to be overprotected while others insufficiently protected. For example, in all the sensitive topics are detected using an absolute metric called surprise based on the information theory, assuming that the interests with less user document support are more sensitive. However, this assumption can be doubted with a simple counter.

Many personalization techniques require iterative user interactions when creating personalized search results. We require to refine the search results with some metrics which require multiple user interactions, such as rank scoring, average rank, and so on. This paradigm is however, infeasible for runtime profiling, as it will not only pose too much risk of privacy breach, but also demand prohibitive processing time for profiling. Thus, we need predictive metrics to measure the search quality and breach risk after personalization, without incurring iterative user interaction.

Disadvantage are as follows:

- All the sensitive topics are detected using an absolute metric called surprisal based on the information theory.
- The existing profile-based PWS do not support runtime profiling.
- The existing methods do not take into account the customization of privacy requirements.
- Personalization techniques require iterative user interactions when creating personalized search results.

2. EXISTING SYSTEM

Personalization techniques require iterative user interactions when creating personalized search results. Below points explain existing behaviors.

On Profile based PWS , The user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly. Profile-based personalization may not even help to improve the search quality for some ad hoc queries, though exposing user profile to a server has put the user's privacy at risk. A better approach is to make an online decision on whether to personalize the query and what to expose in the user profile at runtime.

On Customization of privacy requirements , This considers all the sensitive topics are detected using an absolute metric called surprisal based on the information theory. Assuming that the interests with less user document support are more sensitive.

On Iterative user interactions, They usually refine the search results with some metrics which require multiple user interactions, such as rank scoring, average rank, and so on. This paradigm is, however, infeasible for runtime profiling, as it will not only pose too much risk of privacy breach, but also demand prohibitive processing time for profiling. Thus, we need predictive metrics to measure the search quality, and breach risk after personalization, without incurring iterative user interaction.

3. PROPOSED METHODOLOGY

- We propose UPS (User customizable Privacy-preserving Search) framework is a privacy-preserving personalized web search framework which can generalize profiles for each query according to user-specified privacy requirements.
- We develop two generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. GreedyDP tries to maximize the discriminating power (DP), GreedyIL attempts to minimize the information loss (IL).
- The framework assumes that the queries do not contain any sensitive information, and aims at protecting

the privacy in individual user profiles while retaining their usefulness for PWS.

- UPS consists of a non trusty search engine server and a number of clients. Each user accessing the search service trusts no one but himself or herself.
- The key component for privacy protection is an online profile implemented as a search proxy running on the client machine itself.
- The proxy maintains both the complete user profile, in a hierarchy of nodes with semantics, and the user-specified (customized) privacy requirements represented as a set of sensitive-nodes.
- During the offline phase, a hierarchical user profile is constructed and customized by the user-specified privacy requirements.
- The online phase handles queries as when a user issues a query Qi on the client, the proxy generates a user profile in runtime in the light of query terms. The output of this step is a generalized user profile GI satisfying the privacy requirements. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined in user profiles.
- The query and the generalized user profile are sent together to the PWS server to personalized search.
- The search results are personalized with the profile and delivered back to the query proxy.
- Finally, the proxy either presents the raw results of the use ranks them with the complete user profile.

Taxonomy Repository

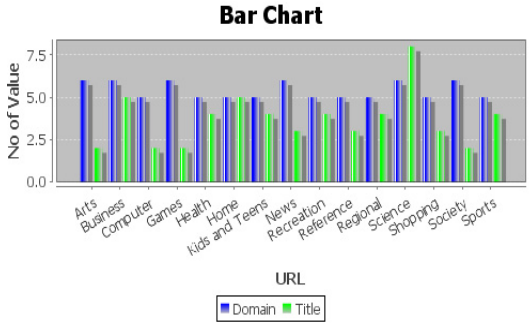

Fig.5 a) Sample Taxonomy Repository

A diagram of a sample user profile is illustrated in which is constructed based on the sample taxonomy repository. We can observe that the owner of this profile is mainly interested in Computer Science and Music, because the major portion of this profile is made up of fragments from taxonomies of these two topics in the sample repository. Some other taxonomies also serve in comprising the profile, for example, Sports and Adults.

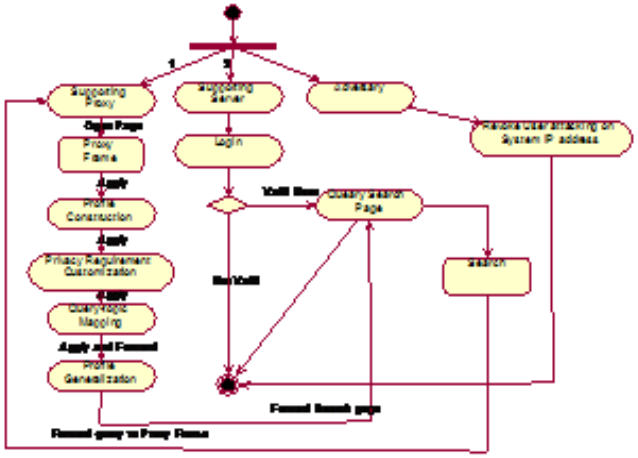

Fig.: 5 b)System flow Chart of Proposed System

Advantages

- UPS provides runtime profiling, which in effect optimizes the personalization utility while respecting user's privacy requirements;
- Allows for customization of privacy needs; and
- Does not require iterative user interaction.
- Provides an inexpensive mechanism for the client to decide whether to personalize a query in UPS.
- It enhances the stability of the search quality.
- It avoids the unnecessary exposure of the user profile.

4. Result And Observations

According to GreedyDP and GreedyIL Algorithm in our system We are presenting graph base result of efficiency of Algorithm in terms of Query Time during for executing each methods.

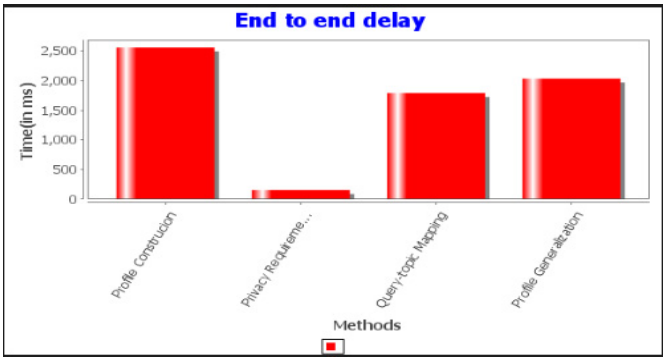

Fig.:-6 a) Methods versus Time

Following performance graph show existing seed profile generalization and green bar showing proposed system. It is showing that it increases according to time but existing system is keep constant.

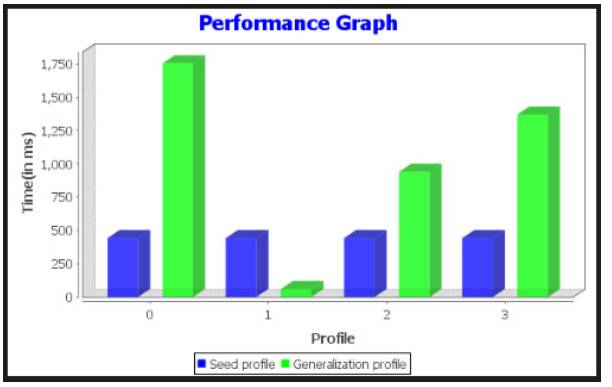

Fig:-6 b) Performance graph of Existing Seed profile and Implementing generalization profile

5. Conclusion

We implemented system a client-side privacy protection framework. System is potentially be adopted by any PWS that captures user profiles in a hierarchical taxonomy. The framework allowed users to specify customized privacy requirements via the hierarchical profiles. In addition, Online generalization on user profiles to protect the personal privacy without compromising the search quality. GreedyDP and GreedyIL algorithm for the online generalization. Our experimental results revealed that system could achieve quality search results while preserving user's customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution.

6. Reference

[1] Lidan Shou, He Bai, Ke Chen, and Gang Chen on "Supporting Privacy Protection, in Personalized WebSearch," Proc. IEEE transactions on knowledge and data engineering VOL:26 NO:2 YEAR 2014.
[2] Jayant Kiran simply. On "Personalize Web search with Enhanced Privacy Protection" on International Journal of Advanced Research in Computer Science and software Engineering.
[3] Ricardo Baeza-Yates, Carlos Hurtado on "Query Recommendation using Query Logs in Search Engine".
[4] Z. Do, R. Song, and J. -R. Wen, on "A Large-Scale Evaluation and Analysis of Personalized Search Strategies," Conf. World Wide Web (WWW), pp. 581-590, 2007.
[5] J. Teevan, S.T. Dumais, and E. Horvitz on "Personalizing Search via Automated Analysis of Interests and Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449-456, 2005.
[6] M. Sparta and S. Gach, on "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
[7] B. Tan, X. Shen, and C. Zhai on "Mining Long-Term Search History to Improve Search Accuracy," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2006.
[8] K. Sugiyama, K. Hatano, and M. Yoshikawa on "Adaptive Web Search Based on User Profile Constructed without any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW), 2004.
[9] Ricardo Baeza-Yates1, Carlos Hurtado1, and Marcelo Mendoza on "Query Recommendation using Query Logs in Search Engines," Proc. Millennium Nucleus, Center for Web Research (P01-029-F), Mideplan, Chile.
[10] Thorsten Joachims, Laura Granka, Bing Pan on "Accurately Interpreting Clickthrough Data as Implicit Feedback," Proc. SIGIR'05, August 15–19, 2005, Salvador, Brazil.
[11] Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, Hang Li on "Context-Aware Query Suggestion by Mining Click-Through and Session Data," KDD'08, August 24–27, 2008, Las Vegas, Nevada, USA.